POINTWISE CONVERGENCE FOR CUBIC AND POLYNOMIAL MULTIPLE ERGODIC AVERAGES OF NON-COMMUTING TRANSFORMATIONS

QING CHU AND NIKOS FRANTZIKINAKIS

Abstract. We study the limiting behavior of multiple ergodic averages involving several not necessarily commuting measure preserving transformations. We work on two types of averages, one that uses iterates along combinatorial parallelepipeds, and another that uses iterates along shifted polynomials. We prove pointwise convergence in both cases, thus answering a question of I. Assani in the former case, and extending results of B. Host-B. Kra and A. Leibman in the latter case. Our argument is based on some elementary uniformity estimates of general bounded sequences, decomposition results in ergodic theory, and equidistribution results on nilmanifolds.

1. Main results

In this paper we study the limiting behavior, in the mean and pointwise, of multiple ergodic averages involving measure preserving transformations that do not necessarily commute. We focus our attention on two such types, special cases of which have previously attracted some attention. One involves iterates taken along combinatorial parallelepipeds, and the other involves iterates taken along shifted polynomials.

1.1. Cubic Averages.

For $k \in \mathbb{N}$ we set

$$V_k := \{0,1\}^k \quad \text{and} \quad V_k^* := V_k \setminus \{\mathbf{0}\}$$

where $\mathbf{0} := (0,0,\cdots,0)$. Let $(X,\mathcal{X},\mu)$ be a probability space[1], and for $\boldsymbol{\epsilon} \in V_k^*$ let $T_{\boldsymbol{\epsilon}} \colon X \to X$ be measure preserving transformations and $f_{\boldsymbol{\epsilon}} \in L^\infty(\mu)$ be functions. We are going to study the limiting behavior of certain multiple ergodic averages taken along k-dimensional combinatorial parallelepipeds of iterates of the transformations $T_{\boldsymbol{\epsilon}}$. More precisely, the *cubic averages of dimension* k are given by

$$A_{k,N}(T_{\boldsymbol{\epsilon}}, f_{\boldsymbol{\epsilon}})(x) := \frac{1}{N^k} \sum_{\mathbf{n}\in[1,N]^k} \prod_{\boldsymbol{\epsilon}\in V_k^*} f_{\boldsymbol{\epsilon}}(T_{\boldsymbol{\epsilon}}^{\boldsymbol{\epsilon}\cdot\mathbf{n}} x) \tag{1}$$

where for $\boldsymbol{\epsilon} = (\epsilon_1,\dots,\epsilon_k) \in V_k$ and $\mathbf{n} = (n_1,\dots,n_k) \in \mathbb{N}^k$ we define

$$\boldsymbol{\epsilon}\cdot\mathbf{n} := \epsilon_1 n_1 + \cdots + \epsilon_k n_k.$$

For instance, the cubic averages of dimension 1 are the ergodic averages, the cubic averages of dimension 2 are defined by

$$\frac{1}{N^2} \sum_{1\le m,n\le N} f_1(T_1^m x)\cdot f_2(T_2^n x)\cdot f_3(T_3^{m+n} x),$$

2000 *Mathematics Subject Classification.* Primary: 37A45; Secondary: 28D05, 11B30.

The second author was partially supported by Marie Curie IRG 248008.

[1]Throughout the text all probability spaces are assumed to be Lebesgue.

and the cubic averages of dimension 3 are similarly defined, using iterates of 7 transformations taken along the combinatorial parallelepipeds $m, n, r, m+n, m+r, n+r, m+n+r$.

The averages $A_{k,N}(T_{\boldsymbol{\epsilon}}, f_{\boldsymbol{\epsilon}})(x)$ are closely linked to the Gowers-Host-Kra seminorms $\lVert\!| \cdot |\!\rVert_k$ that have been used extensively in ergodic theory to find convenient majorants for various other multiple ergodic averages. In [14] it is shown that for ergodic systems $(X, \mathcal{X}, \mu, T)$, and real valued functions $f \in L^\infty(\mu)$, we have

$$\lVert\!| f |\!\rVert_k^{2^k} = \lim_{N\to\infty} \int f \cdot A_{k,N}(T, f)\; d\mu$$

where $A_{k,N}(T, f)$ is defined by letting $T_{\boldsymbol{\epsilon}} = T$ and $f_{\boldsymbol{\epsilon}} = f$ in (1) for every $\boldsymbol{\epsilon} \in V_k^*$. This identity also holds for non-ergodic systems once the seminorms $\lVert\!| \cdot |\!\rVert_k$ are appropriately defined.

The study of the limiting behavior of the averages (1) was initiated by V. Bergelson in [6], where convergence in $L^2(\mu)$ was shown in dimension 2 under the extra assumption that all the transformations are equal. Under the same assumption, Bergelson's result was extended by B. Host and B. Kra for cubic averages of dimension 3 in [13], and for arbitrary dimension k in [14]. More recently in [3], I. Assani established pointwise convergence for cubic averages of arbitrary dimension k when all the transformations are equal. In the same article, and prior to this in [1] and [2], convergence for general, not necessarily commuting transformations, was studied for the first time. Pointwise convergence was established for 2-dimensional averages, and some partial results were obtained for dimensions greater than 2, including convergence when all the transformations are weak mixing. In this article we complete this study by proving pointwise convergence for the cubic averages of arbitrary dimension.

Theorem 1.1. *Let $k \in \mathbb{N}$, $(X, \mathcal{X}, \mu)$ be a probability space, and for $\boldsymbol{\epsilon} \in V_k^*$ let $T_{\boldsymbol{\epsilon}} \colon X \to X$ be measure preserving transformations, and $f_{\boldsymbol{\epsilon}} \in L^\infty(\mu)$ be functions. Then the cubic averages of dimension k, given by* (1)*, converge pointwise as $N \to \infty$.*

It is interesting to contrast the limiting behavior of the cubic averages with some other similar looking averages. To begin with, the averages $\frac{1}{(N-M)^2}\sum_{M<m,n\leq N} f_1(T_1^m x) \cdot f_2(T_2^n x) \cdot f_3(T_3^{m+n} x)$, and their higher dimensional relatives, do not in general converge pointwise (for an example when f_2=f_3=1 see [17]). On the other hand, our argument can be easily adapted to prove convergence in $L^2(\mu)$ for such averages. As for the averages $\frac{1}{N^2}\sum_{1\leq m,n\leq N} f_1(T^n x) \cdot f_2(S^m x) \cdot f_3(T^n S^m x)$, and the "diagonal averages" $\frac{1}{N}\sum_{n=1}^N f_1(T^n x) \cdot f_2(S^n x)$, it is known that they do not converge in general, even in $L^2(\mu)$, unless one makes some commutativity assumption about the transformations T and S (for counterexamples, see [20] for the former, and [4] or [9] for the latter). In fact, even under the assumption that all transformations commute, pointwise convergence of these averages and their higher dimensional relatives is not known.

A key concept that underlies the convergence result of Theorem 1.1 is the *characteristic factors*, meaning a collection of $T_{\boldsymbol{\epsilon}}$-invariant sub-$\sigma$-algebras $\mathcal{Y}_{\boldsymbol{\epsilon}}$, having the property that the difference $A_{k,N}(T_{\boldsymbol{\epsilon}}, f_{\boldsymbol{\epsilon}})(x) - A_{k,N}(T_{\boldsymbol{\epsilon}}, \tilde{f}_{\boldsymbol{\epsilon}})(x)$, where $\tilde{f}_{\boldsymbol{\epsilon}} = \mathbb{E}(f_{\boldsymbol{\epsilon}}|\mathcal{Y}_{\boldsymbol{\epsilon}})$, converges pointwise to 0. Our main goal is to make a suitable choice so that the corresponding factor systems have very special algebraic structure. This is done by controlling the averages $A_{k,N}(T_{\boldsymbol{\epsilon}}, f_{\boldsymbol{\epsilon}})$ by certain seminorms (their precise definiton is given in Section 2.2), thus obtaining the following result:

Theorem 1.2. *Let $k \in \mathbb{N}$, $(X, \mathcal{X}, \mu)$ be a probability space, and for $\boldsymbol{\epsilon} \in V_k^*$ let $T_{\boldsymbol{\epsilon}} \colon X \to X$ be measure preserving transformations, and $f_{\boldsymbol{\epsilon}} \in L^\infty(\mu)$ be functions. Furthermore, suppose that $\lVert\!| f_{\boldsymbol{\epsilon}} |\!\rVert_{k,T_{\boldsymbol{\epsilon}}} = 0$ for some $\boldsymbol{\epsilon} \in V_k^*$. Then the cubic averages of dimension k, given by* (1)*, converge pointwise to* 0 *as $N \to \infty$.*

In fact we give explicit bounds relating the pointwise limiting behavior of the cubic averages (1) and the seminorms $\|\!|f_\epsilon|\!\|_{k,T_\epsilon}$ (see Corollary 3.7).

Using different terminology, Theorem 1.2 states that the factors $\mathcal{Z}_{k-1,T_\epsilon}$, defined in Section 2.4, are characteristic factors for pointwise convergence of the averages (1).

To prove Theorem 1.2 we simplify and extend to our particular context an argument given by Assani in [3]. To prove Theorem 1.1 we combine Theorem 1.2 with the decomposition result of Proposition 3.8 (which was proved in [10] using the structure theorem of [14]). We eventually reduce matters to a known convergence property of nilsequences (all notions are defined in Section 2).

1.2. **Polynomial averages.** We are going to generalize some convergence results of B. Host and B. Kra [15] and A. Leibman [22] that involve multiple ergodic averages of a single transformation to the case that involves several not necessarily commuting transformations.

Theorem 1.3. *Let $\ell \in \mathbb{N}$, and $(X, \mathcal{X}, \mu)$ be a probability space. For $i = 1, \dots, \ell$ let $T_i \colon X \to X$ be measure preserving transformations, $f_i \in L^\infty(\mu)$ be functions, $p_i \in \mathbb{Z}[t]$ be non-constant polynomials such that $p_i - p_j$ is non-constant for $i \neq j$, and $b \colon \mathbb{N} \to \mathbb{N}$ be a sequence such that $b(N) \to \infty$ and $b(N)/N^{1/d} \to 0$ as $N \to \infty$, where d is the maximum degree of the polynomials p_i.*[2] *Then the averages*

$$\frac{1}{Nb(N)} \sum_{1\leq m\leq N, 1\leq n\leq b(N)} f_1(T_1^{m+p_1(n)}x) \cdot \ldots \cdot f_\ell(T_\ell^{m+p_\ell(n)}x) \tag{2}$$

converge pointwise as $N \to \infty$.

Using this result for $\ell + 1$ in place of ℓ, letting $T_0 = \cdots = T_\ell = T$, $p_0 = 0$, and integrating with respect to μ, we deduce that the averages

$$\frac{1}{N} \sum_{n=1}^{N} f_1(T^{p_1(n)}x) \cdot \ldots \cdot f_\ell(T^{p_\ell(n)}x) \tag{3}$$

converge weakly in $L^2(\mu)$ as $N \to \infty$. This recovers one of the main results from [15]. Let us remark though that we were not able to deduce from Theorem 1.3 anything useful regarding the well known open problem of convergence (weakly, in the mean, or pointwise) of the averages $\frac{1}{N}\sum_{n=1}^{N} f_1(T_1^{p_1(n)}x) \cdot \ldots \cdot f_\ell(T_\ell^{p_\ell(n)}x)$ for general commuting measure preserving transformations $T_1, \dots, T_\ell$.

A key ingredient in the proof of Theorem 1.3 is the following result; it plays the same role Theorem 1.2 plays in the proof of Theorem 1.1.

Theorem 1.4. *Under the assumptions of Theorem 1.3, there exists $k \in \mathbb{N}$, depending only on ℓ and the maximum degree of the polynomials $p_1, \dots, p_\ell$, such that: If $\|\!|f_i|\!\|_{k,T_i} = 0$ for some $i \in \{1, \dots, \ell\}$, then the averages*

$$\frac{1}{N} \sum_{m=1}^{N} \Big| \frac{1}{b(N)} \sum_{n=1}^{b(N)} f_1(T_1^{m+p_1(n)}x) \cdot \ldots \cdot f_\ell(T_\ell^{m+p_\ell(n)}x) \Big|^2 \tag{4}$$

converge pointwise to 0 as $N \to \infty$.

---

[2] The second condition guarantees that the contribution of several boundary terms is negligible. For instance, for every bounded sequence $(a(n))_{n\in\mathbb{N}}$ and polynomial $p \in \mathbb{Z}[t]$ with degree at most d, the difference of the averages $\mathbb{E}_{1\leq n\leq b(N), p(n)\leq m\leq N+p(n)}\, a(n)$ and $\mathbb{E}_{1\leq n\leq b(N), 1\leq m\leq N}\, a(n)$ goes to 0 as $N \to \infty$.

It follows at once that the factors $\mathcal{Z}_{k-1,T_{\boldsymbol{\epsilon}}}$, defined in Section 2, are characteristic factors for pointwise convergence of the averages (2) and (4).

Using Theorem 1.4 for $T_1 = \cdots = T_\ell = T$, and integrating with respect to μ, we deduce that there exists $k \in \mathbb{N}$ such that if $|||f_i|||_{k,T} = 0$ for some $i \in \{1, \ldots, \ell\}$, then the averages (3) converge to 0 in $L^2(\mu)$ as $N \to \infty$. This recovers one of the main results from [22] needed to prove convergence in $L^2(\mu)$ for the averages (3).

1.3. **Open problems related to multiple recurrence.** We state some multiple recurrence problems that are naturally related to the previously established convergence results. Historically, recurrence problems have turned out to be easier to establish than the corresponding convergence problems, but this does not seem to be the case in our current setup.

Problem 1. *Let $k \in \mathbb{N}$, $(X, \mathcal{X}, \mu)$ be a probability space, and for $\boldsymbol{\epsilon} \in V_k$ let $T_{\boldsymbol{\epsilon}} \colon X \to X$ be measure preserving transformations. Is it true that for every $A \in \mathcal{X}$ with $\mu(A) > 0$ there exists $\mathbf{n} \in \mathbb{N}^k$ such that*

$$\mu\Big(\bigcap_{\boldsymbol{\epsilon} \in V_k} T_{\boldsymbol{\epsilon}}^{-\boldsymbol{\epsilon}\cdot\mathbf{n}} A\Big) > 0 \ ?$$

We believe that the answer is positive. When all the transformations commute this is indeed the case. Furthermore, the answer is positive when all the transformations are weak mixing since in this case the corresponding averages converge to $(\mu(A))^{2^k}$ (see [3], or use Theorem 1.2 in the current article). In general, even the case $k = 2$ is open. Namely, it is not known whether if T, S, R are measure preserving transformations acting on the same probability space $(X, \mathcal{X}, \mu)$, and $A \in \mathcal{X}$ satisfies $\mu(A) > 0$, then there exist $m, n \in \mathbb{N}$ such that

$$\mu(A \cap T^{-m} A \cap S^{-n} A \cap R^{-m-n} A) > 0. \tag{5}$$

This problem was first studied by Assani in [2]. We remark that using Theorem 1.2, one can reduce matters to verifying this multiple recurrence property for very special systems (namely, systems with ergodic components rotations on compact abelian groups), but we were not able to handle this seemingly simple case. The non-ergodicity of the transformations causes serious problems and another obstacle (that becomes more serious in dimension higher than 2) is that it is not clear why various approximations arguments that one would like to use preserve the recurrence property (5). Interestingly, we were able to overcome the analogous problems for questions pertaining to convergence. Let us also remark that in general no power of $\mu(A)$ can be used as a lower bound for the multiple intersections in (5). To see this let $S = T^{-2}$, $R = T^2$ and factor out the transformation T^{-2n}; then the left hand side in (5) becomes greater than $\mu(A \cap T^{-(m+2n)} A \cap T^{-2(m+2n)} A)$, and it is known that in general no power of $\mu(A)$ can be used as a lower bound for these expressions (see Theorem 2.1 in [7]).

Problem 2. *Let $\ell \in \mathbb{N}$, $(X, \mathcal{X}, \mu)$ be a probability space, and $T_1, \ldots, T_\ell$ be measure preserving transformations acting on X. Furthermore, let $p_1, \ldots, p_\ell$ be distinct polynomials with integer coefficients that satisfy $p_i(0) = 0$ for $i = 1, \ldots, \ell$. Is it true that for every $A \in \mathcal{X}$ with $\mu(A) > 0$ there exist $m, n \in \mathbb{N}$ such that*

$$\mu(A \cap T_1^{-m-p_1(n)} A \cap \cdots \cap T_\ell^{-m-p_\ell(n)} A) > 0 \ ?$$

Again, we believe that the answer is positive. Notice that the case where $T_1 = \cdots = T_\ell$ corresponds to the so called "polynomial Szemerédi Theorem" proved by Bergelson and Leibman [8]. When all transformations are weak mixing the answer is positive since in this case the corresponding averages converge to $(\mu(A))^{\ell+1}$ (this follows from Theorem 1.4). In

general, even the case where all the polynomials are linear is open. Lastly, let us note that the assumption that the polynomials are distinct is necessary. It is known (see for example [9]), that there exist (non-commuting) transformations T, S, acting on the same probability space $(X, \mathcal{X}, \mu)$, and a set $A \in \mathcal{X}$ with $\mu(A) > 0$ and such that $\mu(T^n A \cap S^n A) = 0$ for every $n \in \mathbb{N}$.

1.4. **General conventions and notation.** The following notation will be used throughout the article: $\mathbb{N} := \{1, 2, \dots\}$, $Tf := f \circ T$, $\Re(z)$ is the real part of a complex number z. We write $a\colon \mathbb{Z}_N \to \mathbb{C}$ when $a\colon \mathbb{N} \to \mathbb{C}$ is a periodic sequence with period N. We use boldface symbols for vectors. If F is a finite set and $a\colon F \to \mathbb{C}$, then $\mathbb{E}_{n\in F} a(n) := \frac{1}{|F|}\sum_{n\in F} a(n)$. For $r \in \mathbb{N}$, we denote by $S_r a$ the sequence defined by $(S_r a)(n) := a(n + r)$. We use the symbol $\ll$ when some expression is majorized by a constant multiple of some other expression. If this constant depends on some variables $k_1, \dots, k_\ell$ we write $\ll_{k_1,\dots,k_\ell}$.

2. Background Material

We gather some basic background material that we use throughout this article.

2.1. Basic facts from ergodic theory.

Systems. A *system* is a quadruple $(X, \mathcal{X}, \mu, T)$ where $(X, \mathcal{X}, \mu)$ is a Lebesgue probability space and $T\colon X \to X$ is an invertible measure preserving transformation.

Factors. For the context of this article, a *factor* of a system $(X, \mathcal{X}, \mu, T)$, is a system $(X, \mathcal{Y}, \mu, T)$ where $\mathcal{Y}$ is a T-invariant sub-σ-algebra of $\mathcal{X}$. We often abuse terminology and refer to $\mathcal{Y}$ in place of the quadruple $(X, \mathcal{Y}, \mu, T)$.

Isomorphic systems. Two systems $(X, \mathcal{X}, \mu, T)$ and $(Y, \mathcal{Y}, \nu, S)$ are *isomorphic* if there exists a bijective measurable map $\pi\colon X' \to Y'$, where X' is a T-invariant subset of X and Y' is an S-invariant subset of Y, both of full measure, such that $\mu \circ \pi^{-1} = \nu$ and $(S \circ \pi)(x) = (\pi \circ T)(x)$ for every $x \in X'$.

Ergodicity and ergodic decomposition. We define $\mathcal{I} := \{A \in \mathcal{X}\colon \mu(T^{-1}A \triangle A) = 0\}$. A system is *ergodic* if $\mathcal{I}$ consists only of sets with measure 0 or 1. Given an ergodic system and $f \in L^1(\mu)$, the *ergodic theorem* states that for μ almost every $x \in X$ we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(T^n x) = \int f \, d\mu.$$

Let $x \mapsto \mu_x$ be a regular version of the conditional measures with respect to the σ-algebra $\mathcal{I}$. This means that the map $x \mapsto \mu_x$ is $\mathcal{I}$-measurable, and for every bounded measurable function f we have

$$\mathbb{E}_\mu(f|\mathcal{I})(x) = \int f \, d\mu_x \quad \text{for } \mu \text{ almost every } \ x \in X.$$

Then the *ergodic decomposition* of μ is

$$\mu := \int \mu_x \, d\mu(x). \tag{6}$$

The measures μ_x have the additional property that for μ almost every $x \in X$ the system $(X, \mathcal{X}, \mu_x, T)$ is ergodic.

2.2. **The seminorms $\|\!|\cdot|\!\|_k$.** The seminorms $\|\!|\cdot|\!\|_k$ were defined for ergodic systems in [14]. These definitions can be easily extended to non-ergodic systems.

Given a system $(X,\mathcal{X},\mu,T)$ with ergodic decomposition as in (6) and a function $f\in L^\infty(\mu)$, we define inductively

$$\|\!|f|\!\|_1 := \Big\|\int f\ d\mu_x\Big\|_{L^2(\mu)}\ ; \tag{7}$$

$$\|\!|f|\!\|_{k+1}^{2^{k+1}} := \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\|\!|\bar{f}\cdot T^nf|\!\|_k^{2^k}. \tag{8}$$

It can be shown that for every $k\in\mathbb{N}$ the limit above exists, and $\|\!|\cdot|\!\|_k$, thus defined, is a seminorm on $L^\infty(\mu)$ (see [14], [10]). If further clarification is needed, we write $\|\!|\cdot|\!\|_{k,\mu}$, or $\|\!|\cdot|\!\|_{k,T}$.

More explicitly, when $k\geq 2$, one has

$$\|\!|f|\!\|_k^{2^k} = \lim_{N\to\infty}\mathbb{E}_{n_1\in[1,N]}\cdots\lim_{N\to\infty}\mathbb{E}_{n_{k-1}\in[1,N]}\int\Big|\int\prod_{\epsilon\in V_{k-1}}\mathcal{C}^{|\epsilon|}T^{\epsilon\cdot\mathbf{n}}f\ d\mu_x\Big|^2 d\mu \tag{9}$$

where $\mathbf{n}=(n_1,\dots,n_{k-1})$. It follows that if $\|f\|_{L^\infty(\mu)}\leq 1$, then $\|\!|f|\!\|_k\leq\|f\|_{L^1(\mu)}$ for every $k\in\mathbb{N}$.

For every function $f\in L^\infty(\mu)$ we have

$$\|\!|f|\!\|_{k,\mu}^{2^k} = \int\|\!|f|\!\|_{k,\mu_x}^{2^k}\ d\mu(x).$$

It follows that if $\|\!|f|\!\|_{k,\mu}=0$, then $\|\!|f|\!\|_{k,\mu_x}$ for μ almost every $x\in X$.

2.3. **Nilsystems and nilsequences.** A *nilmanifold* is a homogeneous space $X=G/\Gamma$ where G is a nilpotent Lie group, and Γ is a discrete cocompact subgroup of G. If $G_{k+1}=\{e\}$, where G_k denotes the k-the commutator subgroup of G, we say that X is a *k-step nilmanifold.*

A k-step nilpotent Lie group G acts on G/Γ by left translation, where the translation by a fixed element $a\in G$ is given by $T_a(g\Gamma)=(ag)\Gamma$. By m_X we denote the unique Borel probability measure on X that is invariant under the action of G by left translations (called the *normalized Haar measure*), and by $\mathcal{G}/\Gamma$ we denote the completion of the Borel σ-algebra of G/Γ. Fixing an element $a\in G$, we call the system $(G/\Gamma,\mathcal{G}/\Gamma,m_X,T_a)$ a *k-step nilsystem.*

If $X=G/\Gamma$ is a k-step nilmanifold, $a\in G$, $x\in X$, and $f\in C(X)$, we call the sequence $(f(a^nx))_{n\in\mathbb{N}}$ a *basic k-step nilsequence.* A *k-step nilsequence*, is a uniform limit of *basic k-step nilsequences.*

We are going to use the following result of A. Leibman (see Theorem A in [21]):

Theorem 2.1 ([21])**.** *Let $X=G/\Gamma$ be a nilmanifold, $a_1,\dots,a_\ell\in G$, $f_1\dots,f_\ell\in C(X)$, and $p_1,\dots,p_\ell\colon\mathbb{Z}^d\to\mathbb{Z}$ be polynomials. Then for every Følner sequence $(\Phi_N)_{N\in\mathbb{N}}$ in $\mathbb{Z}^d$ and $x_1,\dots,x_\ell\in X$ the averages*

$$\frac{1}{|\Phi_N|}\sum_{\mathbf{n}\in\Phi_N}f_1(a_1^{p_1(\mathbf{n})}x_1)\cdots f_\ell(a_\ell^{p_\ell(\mathbf{n})}x_\ell)$$

converge as $N\to\infty$.

2.4. **The factors $\mathcal{Z}_k$ and their structure.** Given a system $(X, \mathcal{X}, \mu, T)$, it was shown in [14] (for ergodic systems but the same construction works for general systems) that for every $k \geq 1$, there exists a T-invariant sub-σ-algebra $\mathcal{Z}_{k-1}$ of $\mathcal{X}$ that satisfies

$$\text{for } \ f \in L^\infty(\mu), \ \ \mathbb{E}(f|\mathcal{Z}_{k-1}) = 0 \quad \text{if and only if} \quad |||f|||_{k,T} = 0. \tag{10}$$

The connection between the factors $\mathcal{Z}_k$ of a given system and nilsystems is given by the following structure theorem of Host and Kra:

Theorem 2.2 ([14])**.** *Let $k \in \mathbb{N}$ and $(X, \mathcal{X}, \mu, T)$ be a system with ergodic decomposition as in* (6)*. Then for μ almost every $x \in X$ the system $(X, \mathcal{Z}_k, \mu_x, T)$ is an inverse limit of k-step nilsystems.*

The conclusion in the preceding statement means that for μ almost every $x \in X$ for a given measure μ_x there exists an increasing sequence of T-invariant sub-σ-algebras $(\mathcal{X}_j)_{j\in\mathbb{N}}$ (depending on μ_x), such that $\bigvee_{j\in\mathbb{N}} \mathcal{X}_j = \mathcal{X}$ up to sets of μ_x-measure zero, and each system $(X, \mathcal{X}_j, \mu_x, T)$ is isomorphic to a k-step nilsystem.

We remark that although we do not make explicit use of Theorem 2.2 in this article, it is a key ingredient in the proof of Proposition 3.8 that is crucial for our analysis.

3. Characteristic factors and convergence for cubic averages

3.1. **Characteristic factors for cubic averages.** We are going to prove Theorem 1.2. The main idea is best illustrated by considering the case of cubic averages of dimension 2. Assuming for example that $f_1 \in L^\infty(\mu)$ satisfies $|||f_1|||_{2,\mu,T_1} = 0$, and $f_2, f_3 \in L^\infty(\mu)$, our goal is to establish the pointwise identity

$$\lim_{N\to\infty} |\mathbb{E}_{m,n\in[1,N]} f_1(T_1^m x) \cdot f_2(T_2^n x) \cdot f_3(T_3^{m+n} x)| = 0.$$

It suffices to show that for μ almost every $x \in X$ we have

$$\lim_{N\to\infty} \mathbb{E}_{n\in[1,N]} |\mathbb{E}_{m\in[1,N]} f_1(T_1^m x) \cdot f_3(T_3^{m+n} x)|^2 = 0. \tag{11}$$

Using suitable applications of a variation of van der Corput's fundamental lemma (the precise statement is given in Lemma 3.3) we can show (see Proposition 3.6) that the limit in (11) is bounded by a constant multiple of

$$\lim_{N\to\infty} \mathbb{E}_{n\in[1,N]} |\lim_{N\to\infty} \mathbb{E}_{m\in[1,N]} \bar{f}_1(T_1^m x) \cdot f_1(T_1^{m+n} x)|^2. \tag{12}$$

The ergodic theorem implies that for μ almost every $x \in X$ the last limit is equal to

$$\lim_{N\to\infty} \mathbb{E}_{n\in[1,N]} \Big| \int \bar{f}_1(x) \cdot f_1(T_1^n x) \ d\mu_{x,T_1} \Big|^2 = |||f_1|||^4_{2,\mu_x,T_1}$$

where $\mu = \int \mu_{x,T_1} \, d\mu(x)$ is the ergodic decomposition of the measure μ with respect to T_1. Since $|||f_1|||_{2,\mu,T_1} = 0$ implies that $|||f_1|||_{2,\mu_x,T_1} = 0$ for μ almost every $x \in X$, our goal is established.

Since most of the calculations and estimates do not depend on the dynamical structure of the sequences $(f_{\boldsymbol{\epsilon}}(T_{\boldsymbol{\epsilon}}^n x))_{n\in\mathbb{N}}$ (it is only at the very last step of the argument that we use the pointwise ergodic theorem to take advantage of this extra structure) we work them out for general bounded sequences $(a_{\boldsymbol{\epsilon}}(n))_{n\in\mathbb{N}}$.

Key to our study will be some quantities that control the limiting behavior of the cubic averages (1) when the sequences $(f_{\boldsymbol{\epsilon}}(T_{\boldsymbol{\epsilon}}^n x))_{n\in\mathbb{N}}$ are replaced by general bounded sequences $(a_{\boldsymbol{\epsilon}}(n))_{n\in\mathbb{N}}$. Closely related quantities have been defined by T. Gowers in [11] and by B. Host and B. Kra in [16]. We define these and prove some basic estimates in the next subsections.

3.1.1. *Measures of uniformity.* We remind the reader that when we write $b\colon \mathbb{Z}_N \to \mathbb{C}$ we refer to a periodic sequence $b\colon \mathbb{N} \to \mathbb{C}$ with period N. We say that $a = (a_N)_{N\in\mathbb{N}}$, where $a_N : \mathbb{Z}_N \to \mathbb{C}$, is *uniformly bounded*, if there exists a constant $C \in \mathbb{R}$ such that $|a_N(n)| \leq C$ for every $n \in [1, N]$ and $N \in \mathbb{N}$. For $k \in \mathbb{N}$, $z \in \mathbb{C}$, and $\boldsymbol{\epsilon} \in V_k$, we let $|\boldsymbol{\epsilon}| := \epsilon_1 + \cdots + \epsilon_k$, and $\mathcal{C}^k z := z$ if k is even, and $\mathcal{C}^k z := \bar{z}$ if k is odd.

We let

$$\|a\|_{U_1(\mathbb{N})} := \limsup_{N\to\infty} |\mathbb{E}_{n\in[1,N]} a_N(n)|,$$

and for $k \geq 2$ we define

$$\|a\|_{U_k(\mathbb{N})} := \Big(\limsup_{N\to\infty} \mathbb{E}_{n_1\in[1,N]} \cdots \limsup_{N\to\infty} \mathbb{E}_{n_{k-1}\in[1,N]} \limsup_{N\to\infty} \big|\mathbb{E}_{m\in[1,N]} \prod_{\boldsymbol{\epsilon}\in V_{k-1}} \mathcal{C}^{|\boldsymbol{\epsilon}|} a_N(m + \boldsymbol{\epsilon}\cdot \mathbf{n})\big|^2\Big)^{\frac{1}{2^k}} \tag{13}$$

where $\mathbf{n} = (n_1, \ldots, n_{k-1})$.

Furthermore, for $N \in \mathbb{N}$ we let

$$\|a_N\|_{U_1(\mathbb{Z}_N)} := |\mathbb{E}_{n\in[1,N]} a_N(n)|,$$

and for $k \geq 2$ we define

$$\|a_N\|_{U_k(\mathbb{Z}_N)} := \big(\mathbb{E}_{\mathbf{n}\in[1,N]^{k-1}} \big|\mathbb{E}_{m\in[1,N]} \prod_{\boldsymbol{\epsilon}\in V_{k-1}} \mathcal{C}^{|\boldsymbol{\epsilon}|} a_N(m + \boldsymbol{\epsilon}\cdot \mathbf{n})\big|^2\big)^{\frac{1}{2^k}}.$$

This is the so called Gowers norm of a_N.

Given a bounded sequence $a\colon \mathbb{N} \to \mathbb{C}$, for $N \in \mathbb{N}$, we define $a_N\colon \mathbb{Z}_N \to \mathbb{C}$ by $a_N(n+N\mathbb{Z}) := a(n)$ for $n \in [1, N]$. We let $\tilde{a} := (a_N)_{N\in\mathbb{N}}$. Furthermore, we define

$$\|a\|_{U_k(\mathbb{N})} := \|\tilde{a}\|_{U_k(\mathbb{N})}, \quad \|a\|_{U_k(\mathbb{Z}_N)} := \|a_N\|_{U_k(\mathbb{Z}_N)}.$$

Notice that $\|a\|_{U_k(\mathbb{N})}$ can also be computed by replacing a_N with a in (13).

One immediately sees that $\|\cdot\|_{U_k(\mathbb{N})}$ satisfies the recursive identity

$$\limsup_{N\to\infty} \mathbb{E}_{r\in[1,N]} \|S_r a \cdot \bar{a}\|^{2^k}_{U_k(\mathbb{N})} = \|a\|^{2^{k+1}}_{U_{k+1}(\mathbb{N})}. \tag{14}$$

We caution the reader that the triangle inequality does not necessarily hold for $\|\cdot\|_{U_k(\mathbb{N})}$, but this is not going to play any role in this article.

The next result links the seminorms $\|\cdot\|_{U_k(\mathbb{N})}$ with the ergodic seminorms $\|\cdot\|_k$ that were defined in Section 2.2 (a similar result was also established in Corollary 3.11 of [16]).

Proposition 3.1. *Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving system with ergodic decomposition $\mu = \int \mu_x \, d\mu(x)$ and $f \in L^\infty(\mu)$. Then for μ almost every $x \in X$ we have*

$$\|f(T^n x)\|_{U_k(\mathbb{N})} = \|f\|_{k,\mu_x}.$$

Proof. The ergodic theorem gives that for μ almost every $x \in X$, for every $\mathbf{n} \in \mathbb{N}^{k-1}$ and $\boldsymbol{\epsilon} \in V_{k-1}$ we have

$$\lim_{N\to\infty} \mathbb{E}_{m\in[1,N]} \prod_{\boldsymbol{\epsilon}\in V_{k-1}} \mathcal{C}^{|\boldsymbol{\epsilon}|} f(T^{m+\boldsymbol{\epsilon}\cdot\mathbf{n}} x) = \int \prod_{\boldsymbol{\epsilon}\in V_{k-1}} \mathcal{C}^{|\boldsymbol{\epsilon}|} T^{\boldsymbol{\epsilon}\cdot\mathbf{n}} f \, d\mu_x.$$

The result now follows by using the definition of $\|\cdot\|_{U_k(\mathbb{N})}$ and formula (9). $\square$

3.1.2. *Comparing* $\|\!|\cdot\|\!|_{U_k(\mathbb{Z}_\infty)}$ *with* $\limsup_{N\to\infty}\|\!|\cdot\|\!|_{U_k(\mathbb{Z}_N)}$. The following estimate will be key for our analysis:

Proposition 3.2. *Let* $a=(a_N)_{N\in\mathbb{N}}$, *where* $a_N\colon \mathbb{Z}_N\to\mathbb{C}$, *be uniformly bounded by* 1. *Then for every* $k\in\mathbb{N}$ *we have*

$$\limsup_{N\to\infty}\|\!|a\|\!|_{U_k(\mathbb{Z}_N)}\ll_k \|\!|a\|\!|_{U_k(\mathbb{N})}.$$

To prove Proposition 3.2 we are going to use the following variation of van der Corput's fundamental lemma:

Lemma 3.3. *Let* $N\in\mathbb{N}$ *and* $a\colon\mathbb{Z}_N\to\mathbb{C}$. *Then for every* $R\in\mathbb{N}$ *we have*

$$|\mathbb{E}_{n\in[1,N]}a(n)|^2\leq 2\cdot\mathbb{E}_{r\in[1,R]}\Big(1-\frac{r}{R}\Big)\Re\big(\mathbb{E}_{n\in[1,N]}a(n+r)\cdot\bar{a}(n)\big)+\frac{\mathbb{E}_{n\in[1,N]}|a(n)|^2}{R}.$$

Proof. Let $R\in\mathbb{N}$. Using the identity

$$\mathbb{E}_{n\in[1,N]}a(n)=\mathbb{E}_{n\in[1,N]}\mathbb{E}_{r\in[1,R]}a(n+r)$$

and the Cauchy-Schwarz inequality, we get that $|\mathbb{E}_{n\in[1,N]}a(n)|^2$ is bounded by

$$\mathbb{E}_{n\in[1,N]}|\mathbb{E}_{r\in[1,R]}a(n+r)|^2=\mathbb{E}_{r,r'\in[1,R]}\mathbb{E}_{n\in[1,N]}a(n+r)\cdot\bar{a}(n+r').$$

Isolating those terms for which $r=r'$, and using the symmetry up to conjugation of the remaining expression with respect to r and r', we see that the last expression is equal to

$$\frac{2}{R^2}\cdot\sum_{1\leq r'<r\leq R}\Re\big(\mathbb{E}_{n\in[1,N]}a(n+r)\cdot\bar{a}(n+r')\big)+\frac{\mathbb{E}_{n\in[1,N]}|a(n)|^2}{R}.$$

To end the proof, it suffices to perform the change of variables $n\to n-r'$ and notice that for $k\in\{1,\dots,R\}$ the equation $r-r'=k$ with $1\leq r'<r\leq R$ has $R-k$ solutions. □

Lemma 3.4. *Let* $N\in\mathbb{N}$ *and* $a\colon\mathbb{Z}_N\to\mathbb{C}$ *be bounded by* 1. *Then for every* $R\in\mathbb{N}$ *we have*

$$\mathbb{E}_{n\in[1,N]}|\mathbb{E}_{m\in[1,N]}a(m+n)\cdot\bar{a}(m)|^2\leq 2\cdot\mathbb{E}_{r\in[1,R]}|\mathbb{E}_{m\in[1,N]}a(m+r)\cdot\bar{a}(m)|^2+1/R.$$

Proof. Using Lemma 3.3 we deduce that the left hand side is bounded by

$$2\cdot\mathbb{E}_{n\in[1,N]}\mathbb{E}_{r\in[1,R]}\Big(1-\frac{r}{R}\Big)\Re\big(\mathbb{E}_{m\in[1,N]}a(m+n+r)\cdot\bar{a}(m+r)\cdot\bar{a}(m+n)\cdot a(m)\big)+1/R.$$

Interchanging the averages and performing the change of variables $n\to n-m$ we deduce that the last expression is equal to

$$2\cdot\mathbb{E}_{r\in[1,R]}\Big(1-\frac{r}{R}\Big)|\mathbb{E}_{m\in[1,N]}a(m+r)\cdot\bar{a}(m)|^2+1/R.$$

The result follows. □

Next we prove Proposition 3.2 by successively applying Lemma 3.4.

Proof of Proposition 3.2. Remember that $a_N\colon\mathbb{Z}_N\to\mathbb{C}$ is defined by $a_N(n+N\mathbb{Z}):=a(n)$ for $n\in[1,N]$. For $k=1$ we have

$$\limsup_{N\to\infty}\|\!|a\|\!|_{U_1(\mathbb{Z}_N)}=\limsup_{N\to\infty}\|\!|a_N\|\!|_{U_1(\mathbb{Z}_N)}=\|\!|a\|\!|_{U_1(\mathbb{N})}.$$

Suppose that the statement holds for $k\in\mathbb{N}$, we are going to show that it holds for $k+1$. We have

$$\|\!|a_N\|\!|_{U_{k+1}(\mathbb{Z}_N)}^{2^{k+1}}=\mathbb{E}_{n_1,\dots,n_k\in[1,N]}\Big|\mathbb{E}_{m\in[1,N]}\prod_{\epsilon\in V_k}\mathcal{C}^{|\epsilon|}a_N(m+\epsilon_1n_1+\cdots+\epsilon_kn_k)\Big|^2. \tag{15}$$

We fix $n_1, \dots, n_{k-1} \in [1, N]$, and apply Lemma 3.4 for $A_{N,n_1,\dots,n_{k-1}} \colon \mathbb{Z}_N \to \mathbb{C}$ defined by

$$A_{N,n_1,\dots,n_{k-1}}(m) = \prod_{\boldsymbol{\epsilon}\in V_{k-1}} \mathcal{C}^{|\boldsymbol{\epsilon}|} a_N(m + \epsilon_1 n_1 + \cdots + \epsilon_{k-1} n_{k-1}).$$

We deduce that for every $R, N \in \mathbb{N}$, the right hand side of (15) is bounded by 2 times

$$\mathbb{E}_{n_1,\dots,n_{k-1}\in[1,N]} \mathbb{E}_{n_k\in[1,R]} \Big|\mathbb{E}_{m\in[1,N]} \prod_{\boldsymbol{\epsilon}\in V_k} \mathcal{C}^{|\boldsymbol{\epsilon}|} a_N(m + \epsilon_1 n_1 + \cdots + \epsilon_k n_k)\Big|^2 + 1/R.$$

Next, we fix $n_k \in \mathbb{N}$, and use the inductive hypothesis for the sequence $S_{n_k} a \cdot \bar{a}$. We get

$$\limsup_{N\to\infty} \mathbb{E}_{n_1,\dots,n_{k-1}\in[1,N]} \Big|\mathbb{E}_{m\in[1,N]} \prod_{\boldsymbol{\epsilon}\in V_k} \mathcal{C}^{|\boldsymbol{\epsilon}|} a_N(m + \epsilon_1 n_1 + \cdots + \epsilon_k n_k)\Big|^2 =$$

$$\limsup_{N\to\infty} \|S_{n_k} a \cdot \bar{a}\|_{U_k(\mathbb{Z}_N)}^{2^k} \ll_k \|S_{n_k} a \cdot \bar{a}\|_{U_k(\mathbb{N})}^{2^k}.$$

Combining the previous estimates we get for every positive integer R that

$$\limsup_{N\to\infty} \|a\|_{U_{k+1}(\mathbb{Z}_N)}^{2^{k+1}} = \limsup_{N\to\infty} \|a_N\|_{U_{k+1}(\mathbb{Z}_N)}^{2^{k+1}} \ll_k \mathbb{E}_{n_k\in[1,R]} \|S_{n_k} a \cdot \bar{a}\|_{U_k(\mathbb{N})}^{2^k} + 1/R.$$

Finally, taking the $\limsup$ as $R \to \infty$, and using the identity (14) we get the advertised estimate. □

3.1.3. *Proof of Theorem 1.2.* We first recall a known estimate (see Lemma 3.8 in [11]).

Lemma 3.5 (Gowers-Cauchy-Schwarz Inequality). *Let $k \geq 2$ be an integer, $N \in \mathbb{N}$, and for $\boldsymbol{\epsilon} \in V_{k-1}$ let $a_{\boldsymbol{\epsilon}} \colon \mathbb{Z}_N \to \mathbb{C}$. Then*

$$\mathbb{E}_{\mathbf{n}\in[1,N]^{k-1}} \Big|\mathbb{E}_{m\in[1,N]} \prod_{\boldsymbol{\epsilon}\in V_{k-1}} \mathcal{C}^{|\boldsymbol{\epsilon}|} a_{\boldsymbol{\epsilon}}(m + \boldsymbol{\epsilon}\cdot\mathbf{n})\Big|^2 \leq \prod_{\boldsymbol{\epsilon}\in V_{k-1}} \|a_{\boldsymbol{\epsilon}}\|_{U_k(\mathbb{Z}_N)}^2.$$

Combining Lemma 3.5 and Proposition 3.2 we are going to prove the following key estimate:

Proposition 3.6. *Let $k \geq 2$ be an integer and for $\boldsymbol{\epsilon} \in V_{k-1}$ let $a_{\boldsymbol{\epsilon}} \colon \mathbb{N} \to \mathbb{C}$ be sequences. Then*

$$\limsup_{N\to\infty} \mathbb{E}_{\mathbf{n}\in[1,N]^{k-1}} \Big|\mathbb{E}_{m\in[1,N]} \prod_{\boldsymbol{\epsilon}\in V_{k-1}} \mathcal{C}^{|\boldsymbol{\epsilon}|} a_{\boldsymbol{\epsilon}}(m + \boldsymbol{\epsilon}\cdot\mathbf{n})\Big|^2 \ll_k \prod_{\boldsymbol{\epsilon}\in V_{k-1}} \|a_{\boldsymbol{\epsilon}}\|_{U_k(\mathbb{N})}^2. \tag{16}$$

Proof. We fix $k \geq 2$, $N \in \mathbb{N}$, and for $\boldsymbol{\epsilon} \in V_{k-1}$ we define $a_{\boldsymbol{\epsilon},N} \colon \mathbb{Z}_N \to \mathbb{C}$ as follows:

$$a_{\boldsymbol{\epsilon},N}(n + N\mathbb{Z}) = \begin{cases} a_{\mathbf{0}}(n) \cdot \mathbf{1}_{[1,[N/k]]}(n) & \text{for } \ \boldsymbol{\epsilon} = \mathbf{0} \\ a_{\boldsymbol{\epsilon}}(n) & \text{for } \ \boldsymbol{\epsilon} \neq \mathbf{0} \end{cases}$$

where $\mathbf{0} = (0, 0, \dots, 0)$ and $n \in \{1, \dots, N\}$. Suppose that the element $\mathbf{n} \in \mathbb{N}^{k-1}$ has all its coordinates in the interval $[1, [N/k]]$. Then

$$\prod_{\boldsymbol{\epsilon}\in V_{k-1}} \mathcal{C}^{|\boldsymbol{\epsilon}|} a_{\boldsymbol{\epsilon},N}(m + \boldsymbol{\epsilon}\cdot\mathbf{n}) = \begin{cases} \prod_{\boldsymbol{\epsilon}\in V_{k-1}} \mathcal{C}^{|\boldsymbol{\epsilon}|} a_{\boldsymbol{\epsilon}}(m + \boldsymbol{\epsilon}\cdot\mathbf{n}) & \text{for } \ m \in [1, [N/k]] \\ 0 & \text{for } \ m \in ([N/k], N]. \end{cases}$$

It follows that

$$\mathbb{E}_{\mathbf{n}\in[1,[N/k]]^{k-1}} \Big|\mathbb{E}_{m\in[1,[N/k]]} \prod_{\boldsymbol{\epsilon}\in V_{k-1}} \mathcal{C}^{|\boldsymbol{\epsilon}|} a_{\boldsymbol{\epsilon}}(m + \boldsymbol{\epsilon}\cdot\mathbf{n})\Big|^2$$

is at most

$$k \cdot \mathbb{E}_{\mathbf{n}\in[1,[N/k]]^{k-1}} \Big|\mathbb{E}_{m\in[1,N]} \prod_{\boldsymbol{\epsilon}\in V_{k-1}} \mathcal{C}^{|\boldsymbol{\epsilon}|} a_{\boldsymbol{\epsilon},N}(m + \boldsymbol{\epsilon}\cdot\mathbf{n})\Big|^2,$$

which in turn is at most

$$k^k \cdot \mathbb{E}_{\mathbf{n}\in[1,N]^{k-1}} \Big|\mathbb{E}_{m\in[1,N]} \prod_{\boldsymbol{\epsilon}\in V_{k-1}} \mathcal{C}^{|\boldsymbol{\epsilon}|} a_{\boldsymbol{\epsilon},N}(m+\boldsymbol{\epsilon}\cdot\mathbf{n})\Big|^2.$$

Using Lemma 3.5 we see that the last expression is bounded by a constant multiple of

$$\prod_{\boldsymbol{\epsilon}\in V_{k-1}} \|a_{\boldsymbol{\epsilon},N}\|^2_{U_k(\mathbb{Z}_N)}.$$

Combining the above, taking limits as $N\to\infty$, and using Proposition 3.2, we deduce that the left hand side of (16) is bounded by a constant multiple of

$$\prod_{\boldsymbol{\epsilon}\in V_{k-1}} \|\tilde{a}_{\boldsymbol{\epsilon}}\|^2_{U_k(\mathbb{N})}$$

where $\tilde{a}_{\boldsymbol{\epsilon}} = (a_{\boldsymbol{\epsilon},N})_{N\in\mathbb{N}}$. Furthermore, an easy computation shows that

$$\|\tilde{a}_{\boldsymbol{\epsilon}}\|_{U_k(\mathbb{N})} = \begin{cases} k^{-1}\cdot\|a_{\mathbf{0}}\|_{U_k(\mathbb{N})}, & \text{for } \ \boldsymbol{\epsilon}=\mathbf{0} \\ \|a_{\boldsymbol{\epsilon}}\|_{U_k(\mathbb{N})}, & \text{for } \ \boldsymbol{\epsilon}\neq\mathbf{0}. \end{cases}$$

This completes the proof. □

Applying the previous estimate for suitably chosen sequences we get the following:

Corollary 3.7. *Let $k\geq 2$ be an integer, $(X,\mathcal{X},\mu)$ be a probability space, and for $\boldsymbol{\epsilon}\in V_{k-1}$ let $T_{\boldsymbol{\epsilon}}\colon X\to X$ be measure preserving transformations, and $f_{\boldsymbol{\epsilon}}\in L^\infty(\mu)$ be functions. Furthermore, let $\mu=\int \mu_{x,T_{\boldsymbol{\epsilon}}}\, d\mu(x)$ be the ergodic decomposition of the measure μ with respect to $T_{\boldsymbol{\epsilon}}$. Then for μ almost every $x\in X$ we have*

$$\limsup_{N\to\infty} \mathbb{E}_{\mathbf{n}\in[1,N]^{k-1}} \Big|\mathbb{E}_{m\in[1,N]} \prod_{\boldsymbol{\epsilon}\in V_{k-1}} f_{\boldsymbol{\epsilon}}(T_{\boldsymbol{\epsilon}}^{m+\boldsymbol{\epsilon}\cdot\mathbf{n}}x)\Big|^2 \ll_k \prod_{\boldsymbol{\epsilon}\in V_{k-1}} \|f_{\boldsymbol{\epsilon}}\|^2_{k,\mu_{x,T_{\boldsymbol{\epsilon}}}}.$$

Proof. Let $x\in X$. Applying Proposition 3.6 for the sequences $a_{\boldsymbol{\epsilon}}(n)=f_{\boldsymbol{\epsilon}}(T^n_{\boldsymbol{\epsilon}}x)$, $\boldsymbol{\epsilon}\in V_{k-1}$, we get that the left hand side is bounded by a constant multiple of

$$\prod_{\boldsymbol{\epsilon}\in V_{k-1}} \|f_{\boldsymbol{\epsilon}}(T^n_{\boldsymbol{\epsilon}}x)\|^2_{U_k(\mathbb{N})}.$$

Proposition 3.1 gives that for every $\boldsymbol{\epsilon}\in V_{k-1}$, for μ almost every $x\in X$, we have

$$\|f_{\boldsymbol{\epsilon}}(T^n_{\boldsymbol{\epsilon}}x)\|_{U_k(\mathbb{N})} = \|f_{\boldsymbol{\epsilon}}\|_{k,\mu_{x,T_{\boldsymbol{\epsilon}}}}.$$

This completes the proof. □

We are now one small step from proving Theorem 1.2.

Proof of Theorem 1.2. Suppose that $\|f_{\mathbf{1}}\|_{k,\mu,T_{\mathbf{1}}}=0$, where $\mathbf{1}=(1,0,\cdots,0)$. The proof is similar in the other cases. We want to show that for almost every $x\in X$

$$\lim_{N\to\infty} \mathbb{E}_{\mathbf{n}\in[1,N]^k} \prod_{\boldsymbol{\epsilon}\in V_k^*} f_{\boldsymbol{\epsilon}}(T_{\boldsymbol{\epsilon}}^{\boldsymbol{\epsilon}\cdot\mathbf{n}}x)=0.$$

Using the Cauchy-Schwarz inequality and bounding all functions $f_{(0,\boldsymbol{\epsilon})}$, where $\boldsymbol{\epsilon}\in V_{k-1}$, by their sup norm, we deduce that the expression

$$\Big|\mathbb{E}_{\mathbf{n}\in[1,N]^k} \prod_{\boldsymbol{\epsilon}\in V_k^*} f_{\boldsymbol{\epsilon}}(T_{\boldsymbol{\epsilon}}^{\boldsymbol{\epsilon}\cdot\mathbf{n}}x)\Big|^2$$

is bounded by a constant multiple of an average of the form

$$\mathbb{E}_{\mathbf{n}\in[1,N]^{k-1}}\Big|\mathbb{E}_{m\in[1,N]}\prod_{\boldsymbol{\epsilon}\in V_{k-1}}\tilde{f}_{\boldsymbol{\epsilon}}(\tilde{T}_{\boldsymbol{\epsilon}}^{m+\boldsymbol{\epsilon}\cdot\mathbf{n}}x)\Big|^2, \tag{17}$$

where $\tilde{f}_{\mathbf{1}}=f_{\mathbf{1}}$, $\tilde{f}_{\boldsymbol{\epsilon}}\in\{f_{(1,\boldsymbol{\epsilon})},\boldsymbol{\epsilon}\in V_{k-1}^*\}$, and $\tilde{T}_{\boldsymbol{\epsilon}}\in\{T_{(1,\boldsymbol{\epsilon})},\boldsymbol{\epsilon}\in V_{k-1}^*\}$ for $\boldsymbol{\epsilon}\in V_{k-1}^*$.

Since $|||f_{\mathbf{1}}|||_{k,\mu,T_{\mathbf{1}}}=0$ implies that $|||f_{\mathbf{1}}|||_{k,\mu_x,T_{\mathbf{1}}}=0$ for μ almost every $x\in X$, Corollary 3.7 gives that for μ almost every $x\in X$ the averages (17) converge to 0. This completes the proof. $\square$

3.2. **Convergence of cubic averages.** In this section we are going to prove Theorem 1.1. A natural approach for establishing such a convergence result would be to try to combine Theorem 1.2 with Theorem 2.2, in order to reduce matters to the case where all systems are nilsystems. Such an approach works well when all the transformations are equal, but in our more general setup it presents problems that are difficult to circumvent. For instance, although it is possible to reduce matters to the case where for every $\boldsymbol{\epsilon}\in V_k^*$ the ergodic components of the transformation $T_{\boldsymbol{\epsilon}}$ are inverse limits of nilsystems, the various ergodic disintegrations and sub-σ-algebras involved in the inverse limits cannot be taken to be the same for each transformations $T_{\boldsymbol{\epsilon}}$ (even if the transformations commute). To overcome this problem we work pointwise, and use an approach similar to the one in [10]. We combine Theorem 1.1 with a pointwise decomposition result that applies to general (not necessarily ergodic) systems. It is a direct consequence of Proposition 3.1 from [10] which in turn is a non-trivial consequence of the structure theorem of Host and Kra stated in Theorem 2.2.

Proposition 3.8. *Let $(X,\mathcal{X},\mu,T)$ be a system, $f\in L^\infty(\mu)$, and $k\in\mathbb{N}$. Then for every $\varepsilon>0$, there exist measurable functions f^s,f^u,f^e, with $L^\infty(\mu)$ norm at most $2\,\|f\|_{L^\infty(\mu)}$, such that*

(i) $f=f^s+f^u+f^e$;
(ii) $|||f^u|||_{k+1}=0$; $\|f^e\|_{L^1(\mu)}\le\varepsilon$; *and*
(iii) *for μ almost every $x\in X$, the sequence $(f^s(T^nx))_{n\in\mathbb{N}}$ is a k-step nilsequence.*

Arithmetic versions of this result were recently established in [12] and in [23]. The reader is advised to think of the function f^e as an error term; when one works with convergence problems it typically can be shown to have a negligible effect on our averages (but this is not the case for recurrence problems unless one aims at a uniform lower bound). The function f^u is the uniform component and it too can be neglected once the appropriate uniformity estimates are obtained. Finally, the function f^s is the structured component; this has to be further analyzed, typically using equidistribution results on nilmanifolds.

Proof of Theorem 1.1. For $k=1$ the result follows from the pointwise ergodic theorem. So we can assume that $k\ge2$. Furthermore, we can assume that $\|f_{\boldsymbol{\epsilon}}\|_{L^\infty(\mu)}\le1$ for every $\boldsymbol{\epsilon}\in V_k^*$. Let

$$A_N(f_{\boldsymbol{\epsilon}})(x):=\mathbb{E}_{\mathbf{n}\in[1,N]^k}\prod_{\boldsymbol{\epsilon}\in V_k^*}f_{\boldsymbol{\epsilon}}(T_{\boldsymbol{\epsilon}}^{\boldsymbol{\epsilon}\cdot\mathbf{n}}x).$$

We are going to show that for μ almost every $x\in X$ the sequence $(A_N(f_{\boldsymbol{\epsilon}})(x))_{N\in\mathbb{N}}$ is Cauchy.

By Proposition 3.8, we have that for every $m\in\mathbb{N}$ and $\boldsymbol{\epsilon}\in V_k^*$, there exist measurable functions $f^s_{\boldsymbol{\epsilon},m},f^u_{\boldsymbol{\epsilon},m},f^e_{\boldsymbol{\epsilon},m}$, with $L^\infty(\mu)$ norm bounded by 2, and such that

(i) $f_{\boldsymbol{\epsilon}}=f^s_{\boldsymbol{\epsilon},m}+f^u_{\boldsymbol{\epsilon},m}+f^e_{\boldsymbol{\epsilon},m}$;
(ii) $|||f^u_{\boldsymbol{\epsilon},m}|||_{k,T_{\boldsymbol{\epsilon}}}=0$; $\left\|f^e_{\boldsymbol{\epsilon},m}\right\|_{L^1(\mu)}\le1/m$; and
(iii) for μ almost every $x\in X$, the sequence $(f^s_{\boldsymbol{\epsilon},m}(T_{\boldsymbol{\epsilon}}^nx))_{n\in\mathbb{N}}$ is a $(k-1)$-step nilsequence.

First we study the contribution of the functions $f^u_{\boldsymbol{\epsilon},m}$. Combining property (2) with Theorem 1.2, we see that when evaluating the limit of the averages $A_N(f_{\boldsymbol{\epsilon}})$, we can ignore the contribution of these functions, namely, for every $m \in \mathbb{N}$, for μ almost every $x \in X$ we have

$$\lim_{N\to\infty} |A_N(f_{\boldsymbol{\epsilon}})(x) - A_N(f^s_{\boldsymbol{\epsilon},m} + f^e_{\boldsymbol{\epsilon},m})(x)| = 0. \tag{18}$$

Next, we study the contribution of the functions $f^e_{\boldsymbol{\epsilon},m}$. We are going to show that this too is essentially negligible, as long as we consider suitably large values of m. Indeed, if we expand the expression $A_N(f^s_{\boldsymbol{\epsilon},m} + f^e_{\boldsymbol{\epsilon},m}) - A_N(f^s_{\boldsymbol{\epsilon},m})$, use Corollary 3.7 to bound each of the terms, and also use that $\|\!|f^e_{\boldsymbol{\epsilon},m}\|\!|_{k,\mu_{x,T_{\boldsymbol{\epsilon}}}} \leq 2$ and $\|\!|f^s_{\boldsymbol{\epsilon},m}\|\!|_{k,\mu_{x,T_{\boldsymbol{\epsilon}}}} \leq 2$, we get for μ almost every $x \in X$ the bound

$$\limsup_{N\to\infty} |A_N(f^s_{\boldsymbol{\epsilon},m} + f^e_{\boldsymbol{\epsilon},m})(x) - A_N(f^s_{\boldsymbol{\epsilon},m})(x)| \ll_k \max_{\boldsymbol{\epsilon}\in V_k^*} \|\!|f^e_{\boldsymbol{\epsilon},m}\|\!|_{k,\mu_{x,T_{\boldsymbol{\epsilon}}}} \ll \max_{\boldsymbol{\epsilon}\in V_k^*} \left\|f^e_{\boldsymbol{\epsilon},m}\right\|_{L^1(\mu_{x,T_{\boldsymbol{\epsilon}}})}.$$

By property (ii) we have $\lim_{m\to\infty} \int |f^e_{\boldsymbol{\epsilon},m}| \, d\mu = 0$ for $\boldsymbol{\epsilon} \in V_k^*$., and as a consequence there exists a sequence $(m_l)_{l\in\mathbb{N}}$, with $m_l \to \infty$, and such that for μ almost every $x \in X$ we have

$$\lim_{l\to\infty} \int |f^e_{\boldsymbol{\epsilon},m_l}| \, d\mu_{x,T_{\boldsymbol{\epsilon}}} = 0$$

for every $\boldsymbol{\epsilon} \in V_k^*$. From the preceding discussion it follows that for μ almost every $x \in X$

$$\lim_{l\to\infty} \limsup_{N\to\infty} |A_N(f^s_{\boldsymbol{\epsilon},m_l} + f^e_{\boldsymbol{\epsilon},m_l})(x) - A_N(f^s_{\boldsymbol{\epsilon},m_l})(x)| = 0. \tag{19}$$

Combining (18) and (19) we get for μ almost every $x \in X$ that

$$\lim_{l\to\infty} \limsup_{N\to\infty} |A_N(f_{\boldsymbol{\epsilon}})(x) - A_N(f^s_{\boldsymbol{\epsilon},m_l})(x)| = 0. \tag{20}$$

Since by property (iii), for μ almost every $x \in X$, for every $l \in \mathbb{N}$, and $\boldsymbol{\epsilon} \in V_k^*$, the sequence $(f^s_{\boldsymbol{\epsilon},m_l}(T^n_{\boldsymbol{\epsilon}} x))_{n\in\mathbb{N}}$ is a nilsequence, it follows from Theorem 2.1 that for μ almost every $x \in X$, for every $l \in \mathbb{N}$, the averages $A_N(f^s_{\boldsymbol{\epsilon},m_l})(x)$ converge. Combining this with (20), we deduce that for μ almost every $x \in X$ the sequence $(A_N(f_{\boldsymbol{\epsilon}})(x))_{N\in\mathbb{N}}$ is Cauchy. This completes the proof. □

4. Characteristic factors and convergence for polynomial averages

In this section we are going to prove Theorems 1.3 and 1.4. As was the case with the cubic averages, some uniformity estimates for general bounded sequences play a key role in the argument. We start with establishing these.

4.1. **Uniformity estimates.**

We remind the reader that in the forthcoming statements $b\colon \mathbb{N} \to \mathbb{N}$ is a sequence that satisfies

$$b(N) \to \infty \quad \text{ and } \quad b(N)/N^{1/d} \to 0$$

where d is the maximum degree of the polynomials involved in each statement. To avoid confusion, let us also remark that none of the sequences defined defined in this section is assumed to be periodic.

Our goal in this section is to establish the following estimate:

Proposition 4.1. *Let $a_1, \ldots, a_\ell \colon \mathbb{N} \to \mathbb{C}$ be bounded sequences and $p_1, \ldots, p_\ell \in \mathbb{Z}[t]$ be non-constant polynomials such that $p_i - p_j$ is non-constant for $i \neq j$. Then there exists $k \in$*

$\mathbb{N}$, depending only on ℓ and the maximum degree of the polynomials $p_1,\ldots,p_\ell$, such that if $\|a_i\|_{U_k(\mathbb{N})}=0$ for some $i\in\{1,\ldots,\ell\}$, then

$$\lim_{N\to\infty}\mathbb{E}_{m\in[1,N]}\Big|\mathbb{E}_{n\in[1,b(N)]}\prod_{i=1}^{\ell}a_i(m+p_i(n))\Big|^2=0.$$

It will be more convenient for us to prove a somewhat more involved statement, where the uniform sequence is associated with the polynomial of maximal degree:

Proposition 4.2. *Let $a_1\colon\mathbb{N}\to\mathbb{C}$ be a bounded sequence and $a_{2,N},\ldots,a_{\ell,N}\colon\mathbb{N}\to\mathbb{C}$, $N\in\mathbb{N}$, be a collection of uniformly bounded sequences. Furthermore, let $p_1,\ldots,p_\ell\in\mathbb{Z}[t]$ be polynomials such that p_1-p_i is non-constant for $i=2,\ldots,\ell$, and suppose that $\deg(p_1)\geq\deg(p_i)$ for $i=2,\ldots,\ell$. Then there exists $k\in\mathbb{N}$, depending only on ℓ and on $\deg(p_1)$, such that if $\|a_1\|_{U_k(\mathbb{N})}=0$, then*

$$\lim_{N\to\infty}\mathbb{E}_{m\in[1,N]}\Big|\mathbb{E}_{n\in[1,b(N)]}\Big(a_1(m+p_1(n))\cdot\prod_{i=2}^{\ell}a_{i,N}(m+p_i(n))\Big)\Big|^2=0. \tag{21}$$

Proof of Proposition 4.1 assuming Proposition 4.2. Let $\{p_1,\ldots,p_\ell\}$ be a family of polynomials that satisfies the assumptions of Proposition 4.1. Because of the symmetry of the statement of Proposition 4.1 it suffices to establish its conclusion when $i=1$.

For $N\in\mathbb{N}$ we define a sequence $a_{0,N}\colon\mathbb{N}\to\mathbb{C}$ by

$$a_{0,N}(m):=\mathbb{E}_{n\in[1,b(N)]}\prod_{i=1}^{\ell}\bar{a}_i(m+p_i(n)).$$

Then

$$\mathbb{E}_{m\in[1,N]}\Big|\mathbb{E}_{n\in[1,b(N)]}\prod_{i=1}^{\ell}a_i(m+p_i(n))\Big|^2=\mathbb{E}_{m\in[1,N]}\mathbb{E}_{n\in[1,b(N)]}\Big(a_{0,N}(m)\prod_{i=1}^{\ell}a_i(m+p_i(n))\Big). \tag{22}$$

Let $p\in\{p_1,\ldots,p_\ell\}$ be any polynomial such that the polynomial $p+p_1$ has maximal degree within the family $\{p,p+p_1,\ldots,p+p_\ell\}$. Making the change of variables $m\to m+p(n)$, and using our growth assumption $p(b(N))/N\to 0$, we see that the difference of the averages

$$\mathbb{E}_{n\in[1,b(N)]}\mathbb{E}_{m\in[1,N]}\Big(a_{0,N}(m)\prod_{i=1}^{\ell}a_i(m+p_i(n))\Big) \tag{23}$$

and the averages

$$\mathbb{E}_{n\in[1,b(N)]}\mathbb{E}_{m\in[1,N]}\Big(a_{0,N}(m+p(n))\prod_{i=1}^{\ell}a_i(m+p(n)+p_i(n))\Big) \tag{24}$$

converges to 0 as $N\to\infty$. Since by assumption the polynomials p_1 and p_1-p_i are non-constant for $i=2,\ldots,\ell$, and by the choice of p the polynomial $p+p_1$ has maximal degree within the family $\{p,p+p_1,\ldots,p+p_\ell\}$, the assumptions of Proposition 4.2 are satisfied, where the role of p_1 plays the polynomial $p+p_1$. Using the Cauchy-Schwarz inequality we conclude that there exists $k\in\mathbb{N}$, depending only on ℓ and on $\deg(p+p_1)$, such that if $\|a_1\|_{U_k(\mathbb{N})}=0$, then the averages (24) converge to 0 as $N\to\infty$. As a consequence, the averages (23) converge to 0 as $N\to\infty$. The result now follows from (22). □

We are going to prove Proposition 4.2 by repeated applications of the following consequence of van der Corput's fundamental estimate (see, for example, Lemma 3.1 in [18]):

Corollary 4.3. *Let $N \in \mathbb{N}$ and $a(1), \ldots, a(N)$ be complex numbers bounded by 1. Then for every integer R between 1 and N we have*

$$\left|\mathbb{E}_{n\in[1,N]}a(n)\right|^2 \leq 4 \cdot \Big(\mathbb{E}_{r\in[1,R]}(1-rR^{-1})\Re\big(\mathbb{E}_{n\in[1,N]}a(n+r)\cdot\bar{a}(n)\big) + R^{-1} + RN^{-1}\Big).$$

4.1.1. *The linear case.* The next lemma will be used to prove the linear case of Proposition 4.1. Furthermore, its proof contains the main technical maneuver needed to carry out the inductive step in the proof of Proposition 4.1.

Lemma 4.4. *Let $a\colon \mathbb{N} \to \mathbb{C}$ be a sequence bounded by 1. Then*

$$\limsup_{N\to\infty} \mathbb{E}_{m\in[1,N]}|\mathbb{E}_{n\in[1,b(N)]}a(m+n)|^2 \ll \|a\|_{U_2(\mathbb{N})}.$$

Proof. Since $b(N) \to \infty$ as $N \to \infty$, by Corollary 4.3 we get that for every $R \in \mathbb{N}$ the limit

$$\limsup_{N\to\infty} \mathbb{E}_{m\in[1,N]}|\mathbb{E}_{n\in[1,b(N)]}a(m+n)|^2$$

is bounded by 4 times the expression

$$\limsup_{N\to\infty} \mathbb{E}_{m\in[1,N]}\mathbb{E}_{r\in[1,R]}(1-rR^{-1})\Re\big(\mathbb{E}_{n\in[1,b(N)]}a(m+n+r)\cdot\bar{a}(m+n)\big) + R^{-1}.$$

We interchange averages and make the change of variables $m \to m-n$. Since $b(N)/N \to 0$ and the sequence $(a(n))_{n\in\mathbb{N}}$ is bounded, we deduce that the last expression is equal to

$$\limsup_{N\to\infty} \mathbb{E}_{r\in[1,R]}(1-rR^{-1})\Re\big(\mathbb{E}_{m\in[1,N]}a\big(m+r\big)\cdot\bar{a}(m)\big) + R^{-1}.$$

Finally, letting $R \to \infty$ we get that the original limit is bounded by

$$\limsup_{N\to\infty} \mathbb{E}_{r\in[1,N]} \limsup_{N\to\infty} |\mathbb{E}_{m\in[1,N]}a(m+r)\cdot\bar{a}(m)| \leq \|a\|^2_{U_2(\mathbb{N})}.$$

Since $\|a\|_{U_2(\mathbb{N})} \leq 1$, this establishes the advertised estimate. □

Proof of Proposition 4.2 for linear polynomials. For notational convenience we let $a_{1,N} := a_1$ for every $N \in \mathbb{N}$. It suffices to show that if $\|a_{i,N}\|_\infty \leq 1$ for $i = 1, \ldots, \ell$ and $N \in \mathbb{N}$, then

$$\limsup_{N\to\infty} \mathbb{E}_{m\in[1,N]}\Big|\mathbb{E}_{n\in[1,b(N)]}\prod_{i=1}^{\ell} a_{i,N}(m+k_i n)\Big|^2 \ll_{k_1,\ldots,k_\ell} \|a_1\|_{U_{\ell+1}(\mathbb{N})}. \tag{25}$$

We use induction on ℓ, the number of sequences involved.

For $\ell = 1$ the result follows from Lemma 4.4, and the estimate

$$\limsup_{N\to\infty} \mathbb{E}_{n\in[1,N]}|a(kn)| \leq k \cdot \limsup_{N\to\infty} \mathbb{E}_{n\in[1,N]}|a(n)|. \tag{26}$$

To carry out the inductive step, let $\ell \geq 2$, and suppose that the statement holds for $\ell - 1$ sequences. Following the argument used in the proof of Lemma 4.4, using the induction hypothesis, and the estimate (26), we get that the left hand side in (25) is bounded by a constant, that depends on $k_1, \ldots, k_\ell$, multiple of

$$\limsup_{N\to\infty} \mathbb{E}_{r\in[1,N]}\|S_r a_1 \cdot \bar{a}_1\|_{U_\ell(\mathbb{N})} \leq \|a_1\|^2_{U_{\ell+1}(\mathbb{N})}$$

where the last estimate follows from (14) and Hölder's inequality. Since $\|a_1\|_\infty \leq 1$, we have $\|a_1\|_{U_{\ell+1}(\mathbb{N})} \leq 1$. This completes the proof. □

4.1.2. *The general case.* We first explain an induction scheme, often called PET induction (Polynomial Exhaustion Technique), on types of families of polynomials that was introduced by Bergelson in [5].

We define the *degree* of a family $\mathcal{P}$ of non-constant polynomials to be the maximum of the degrees of the polynomials in the family. Let $\mathcal{P}_i$ be the subfamily of polynomials of degree i in $\mathcal{P}$. We let w_i denote the number of distinct leading coefficients that appear in the family $\mathcal{P}_i$. The vector $(d, w_d, \ldots, w_1)$ is called the *type* of the family of polynomials $\mathcal{P}$. We order the set of all possible types lexicographically, meaning, $(d, w_d, \ldots, w_1) > (d', w'_{d'}, \ldots, w'_1)$ if and only if in the first instance where the two vectors disagree the coordinate of the first vector is greater than the coordinate of the second vector. One easily verifies that every decreasing sequence of types is eventually constant, thus, if some operation reduces the type, then after a finite number of repetitions it is going to terminate.

Next, we define such an operation: Let $\mathcal{P} = (p_1, \ldots, p_\ell)$ be an ordered family of polynomials, $p \in \mathcal{P}$, and $r \in \mathbb{N}$. The family (p, r)-vdC$(\mathcal{P})$ consists of all non-constant polynomials of the form $p_i - p$, $S_r p_i - p$, $i = 1, \ldots, \ell$, where $S_r p$ is defined by $(S_r p)(n) := p(n + r)$. We order them so that the polynomial $S_r p_1 - p$ appears first.

We call an ordered family of polynomials $(p_1, \ldots, p_\ell)$ *nice* if $\deg(p_1) \geq \deg(p_i)$ and $p_1 - p_i$ is non-constant for $i = 2, \ldots, \ell$.

Lemma 4.5. *Let $\mathcal{P} = (p_1, \ldots, p_\ell)$ be a nice ordered family of polynomials, and suppose that $deg(p_1) \geq 2$. Then there exists a polynomial $p \in \mathcal{P}$, such that for every large enough $r \in \mathbb{N}$, the family (p, r)-vdC$(\mathcal{P})$ is nice and has strictly smaller type than that of $\mathcal{P}$.*

Proof. If all the polynomials have the same degree and leading coefficient, then we take $p = p_1$. If all the polynomials have the same degree and at least one has different leading coefficient than p_1, then we take any such polynomial as p. Otherwise, there exists a non-constant polynomial in $\mathcal{P}$ with degree strictly smaller than the degree of p_1. We take p to be any such polynomial that has minimal degree. In all cases, it is easy to check the advertised property. □

Proof of Proposition 4.2. It suffices to show that the k given in the statement of Proposition 4.2 depends only on the number ℓ and the type W of the family of polynomials involved. This is the case because if we fix the degree and the cardinality of a family of polynomials, then there are a finite number of possibilities for its type.

We are going to use induction on the type of the family of polynomials involved. As our base case we take the case where all the polynomials are linear; then the result was proved in the previous subsection with $k = \ell + 1$.

Let now $\mathcal{P}$ be a nice ordered family of ℓ polynomials with $\deg(p_1) \geq 2$ and type W, and suppose that the statement holds for all nice ordered families of ℓ' polynomials with type W' strictly smaller than W for some $k = k(W', \ell') \in \mathbb{N}$.

Let $p \in \mathcal{P}$ be chosen as in Lemma 4.5. Using Corollary 4.3, making the change of variables $m \to m - p(n)$, and using that $p(b(N))/N \to 0$, exactly as in the proof of Lemma 4.4, we get that the limsup as $N \to \infty$ of the averages in (21) is bounded by a constant multiple of

$$\limsup_{N\to\infty} \mathbb{E}_{r\in[1,N]} \limsup_{N\to\infty} \mathbb{E}_{m\in[1,N]} \big|\mathbb{E}_{n\in[1,b(N)]} \prod_{i=1}^{\ell} \bar{a}_{i,N}\big(m+p_i(n+r)-p(n)\big)\cdot a_{i,N}\big(m+p_i(n)-p(n)\big)\big|,$$

where again for notational convenience we have defined $a_{1,N} := a_1$ for $N \in \mathbb{N}$. By Lemma 4.5, for suitably large $r \in \mathbb{N}$, the family (p, r)-vdC$(\mathcal{P})$ is nice, has type strictly smaller than W,

and consists of at most 2ℓ polynomials. Let

$$k(W,\ell) = \max_{W'<W, \ell'\leq 2\ell} k(W',\ell'),$$

where the maximum is taken over all ℓ' with $\ell' \leq 2\ell$ and possible types W' with $W' < W$ of families consisting of at most 2ℓ polynomials (there is a finite number of such possible types). Using the induction hypothesis and the Cauchy-Schwarz inequality, we get that if $\|a_1\|_{U_{k(W,\ell)}(\mathbb{N})} = 0$, then for every large enough $r \in \mathbb{N}$ we have

$$\limsup_{N\to\infty} \mathbb{E}_{m\in[1,N]} \Big| \mathbb{E}_{n\in[1,b(N)]} \Big(\prod_{i=1}^{\ell} \bar{a}_{i,N}\big(m + p_i(n+r) - p(n)\big) \cdot a_{i,N}\big(m + p_i(n) - p(n)\big) \Big) \Big| = 0.$$

This completes the induction and the proof. □

4.2. **Proof of the main results for polynomial averages.** We are now one short step from proving Theorems 1.3 and 1.4.

Proof of Theorem 1.4. Let $k \in \mathbb{N}$ be such that the conclusion of Proposition 4.1 holds. Without loss of generality we can assume that $\|f_1\|_{k,\mu,T_1} = 0$. Then for μ almost every $x \in X$ we have $\|f_1\|_{k,\mu_x,T_1} = 0$. Using Proposition 3.1 we deduce that for μ almost every $x \in X$ we have $\|f_1(T_1^n x)\|_{U_k(\mathbb{N})} = 0$. The result now follows by applying Proposition 4.1 to the sequences $a_i \colon \mathbb{N} \to \mathbb{C}$ defined by $a_i(n) := f_i(T_i^n x)$, $i = 1, \ldots, \ell$. □

Proof of Theorem 1.3. We assume as we may that $\|f_i\|_{L^\infty(\mu)} \leq 1$ for $i = 1, \ldots, \ell$. Let $k \in \mathbb{N}$ be the integer given by Theorem 1.4. Let $\varepsilon > 0$. For $i = 1 \ldots, \ell$, we use Proposition 3.8 to get the decomposition $f_i = f_{i,\varepsilon}^s + f_{i,\varepsilon}^u + f_{i,\varepsilon}^e$, where $\|f_{i,\varepsilon}^u\|_k = 0$, $\left\|f_{i,\varepsilon}^e\right\|_{L^1(\mu)} \leq \varepsilon$, all functions are bounded by 2, and for μ almost every $x \in X$, the sequence $(f_{i,\varepsilon}^s(T^n x))_{n\in\mathbb{N}}$ is a $(k-1)$-step nilsequence. Let

$$A_N(f_i)(x) := \mathbb{E}_{m\in[1,N], n\in[1,b(N)]} f_1(T_1^{m+p_1(n)} x) \cdot \ldots \cdot f_\ell(T_\ell^{m+p_\ell(n)} x).$$

Theorem 2.1 implies that the averages $A_N(f_{i,\varepsilon}^s)(x)$ convergence pointwise. Hence, it suffices to show that when computing the average $A_N(f_i)(x)$ the contribution of the functions $f_{i,\varepsilon}^u$ and $f_{i,\varepsilon}^e$ becomes negligible as $N \to \infty$ and ε is taken suitably small. Theorem 1.4 implies that the contribution of the functions $f_{i,\varepsilon}^u$ is negligible, independently of the choice of ε. To handle the contribution of the functions $f_{i,k}^e$ we argue as in the proof of the corresponding convergence result for the cubic averages in Section 3.2. Let us just explain the only point where our argument deviates slightly from the aforementioned argument. We expand $A_N(f_{i,\varepsilon}^s + f_{i,\varepsilon}^e)$ and write $A_N(f_{i,\varepsilon}^s + f_{i,\varepsilon}^e) - A_N(f_{i,\varepsilon}^s)$ as a sum of $2^\ell - 1$ averages. We deal with each such average separately, and bound all the functions by their sup norm except one (chosen arbitrarily) that is equal to $f_{i,\varepsilon}^e$ for some $i \in \{1, \ldots, \ell\}$. Upon doing this, we get the bound

$$|A_N(f_{i,\varepsilon}^s + f_{i,\varepsilon}^e)(x) - A_N(f_{i,\varepsilon}^s)(x)| \ll_\ell \max_{i=1,\ldots,\ell} \mathbb{E}_{m\in[1,N], n\in[1,b(N)]} |f_{i,\varepsilon}^e|(T_i^{m+p_i(n)} x)$$

That the right hand side becomes negligible as $N \to \infty$, and ε is chosen suitably small, follows (as in the proof given in Section 3.2) upon noticing that for every system $(X, \mathcal{X}, \mu, T)$, function $f \in L^\infty(\mu)$, and polynomial $p \in \{p_1, \ldots, p_\ell\}$, one has for μ almsot every $x \in X$ that

$$\lim_{N\to\infty} \mathbb{E}_{m\in[1,N], n\in[1,b(N)]} |f|(T^{m+p(n)} x) = \int |f| \, d\mu_x$$

where $\mu = \int \mu_x \, d\mu(x)$ is the ergodic decomposition of the measure μ with respect to T. To get this identity it suffices to make the change of variables $m \to m - p(n)$, use that $p(b(N))/N \to 0$ and the ergodic theorem. This completes the proof. $\square$

References

[1] I. Assani. Pointwise convergence along cubes for measure preserving systems. Unpublished manuscript 2003. Available at arXiv:math/0311274
[2] I. Assani. Averages along cubes for not necessarily commuting m.p.t. Ergodic theory and related fields. *Contemp. Math.* **430**, Amer. Math. Soc., Providence, RI, (2007), 1–19.
[3] I. Assani. Pointwise convergence of ergodic averages along cubes. *J. Analyse Math.* **110** (2010), 241–269.
[4] B. Berend. Joint ergodicity and mixing. *J. Analyse Math.* **45** (1985), 255–284.
[5] V. Bergelson. Weakly mixing PET. *Ergodic Theory Dynam. Systems* **7** (1987), no. 3, 337–349.
[6] V. Bergelson. The multifarious Poincare recurrence theorem. Descriptive set theory and dynamical systems. *London Math. Soc. Lecture Note Series* **277**, Cambridge Univ. Press, Cambridge, (2000), 31–57.
[7] V. Bergelson, B. Host, B. Kra, with an Appendix by I. Ruzsa. Multiple recurrence and nilsequences. *Inventiones Math.* **160** (2005), 261-303.
[8] V. Bergelson, A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *Journal Amer. Math. Soc.* **9** (1996), 725–753.
[9] V. Bergelson, A. Leibman. Failure of Roth theorem for solvable groups of exponential growth. *Ergodic Theory Dynam. Systems* **24** (2004), no. 1, 45–53.
[10] Q. Chu, N. Franzikinakis, B. Host. Ergodic averages of commuting transformations with distinct degree polynomial iterates. To appear in *Proc. Lond. Math. Soc.* Available at arXiv:0912.2641
[11] W. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.* **11** (2001), 465–588.
[12] B. Green, T. Tao. An airthmetic regularity lemma, an associated counting lemma, and applications. Preprint 2010. Available at arXiv:1002.2028
[13] B. Host, B. Kra. Averaging along cubes. Modern dynamical systems and applications. *Cambridge Univ. Press*, Cambridge, (2004), 123–144.
[14] B. Host, B. Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. Math.* **161** (2005), 397-488.
[15] B. Host, B. Kra. Convergence of polynomial ergodic averages. *Isr. J. Math.* **149** (2005), 1–19.
[16] B. Host, B. Kra. Uniformity seminorms on l^∞ and spplications. *J. Analyse Math.* **108** (2009), 219–276.
[17] A. del Junco, J. Rosenblatt. Counterexamples in ergodic theory and number theory. *Math. Ann.* **245** (1979), no. 3, 185–197.
[18] L. Kuipers, H. Niederreiter. Uniform distribution of sequences. Pure and Applied Mathematics. *Wiley-Interscience*, New York-London-Sydney, (1974).
[19] A. Leibman. Multiple recurrence theorem for measure preserving actions of a nilpotent group. *Geom. Funct. Anal.* **8** (1998), 853–931.
[20] A. Leibman. Lower bounds for ergodic averages. *Ergodic Theory Dynam. Systems* **22** (2002), no. 3, 863–872.
[21] A. Leibman. Pointwise convergence of ergodic averages for polynomial actions of $\mathbb{Z}^d$ by translations on a nilmanifold. *Ergodic Theory Dynam. Systems* **25** (2005), no. 1, 215–225.
[22] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables. *Isr. J. Math.* **146** (2005), 303–316.
[23] B. Szegedy. Gowers norms, regularization and limits of functions on abelian groups. Preprint 2010. Available at arXiv:1010.6211

(Qing Chu) Université Paris-Est Marne-la-Vallée, Laboratoire d'analyse et de mathématiques appliquées, UMR CNRS 8050, 5 Bd Descartes, 77454 Marne la Vallée Cedex, France
E-mail address: qing.chu@univ-mlv.fr

(Nikos Frantzikinakis) University of Crete, Department of mathematics, Knossos Avenue, Heraklion 71409, Greece
E-mail address: frantzikinakis@gmail.com